



Dudding, T., Johansson, M., Thomas, S. J., Brennan, P., Martin, R. M., & Timpson, N. J. (2018). Assessing the causal association between 25-hydroxyvitamin D and the risk of oral and oropharyngeal cancer using Mendelian randomization. *International Journal of Cancer*, 143(5), 1029-1036. <https://doi.org/10.1002/ijc.31377>

Publisher's PDF, also known as Version of record

License (if available):  
CC BY

Link to published version (if available):  
[10.1002/ijc.31377](https://doi.org/10.1002/ijc.31377)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the final published version of the article (version of record). It first appeared online via WILEY at <https://onlinelibrary.wiley.com/doi/abs/10.1002/ijc.31377>. Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>



# Supporting your research with our capabilities

BD Accuri™ C6 Plus Personal Flow Cytometer

BD FACSCelesta™ Cell Analyzer

BD LSRFortessa™ X-20 Cell Analyzer

BD FACSMelody™ Cell Sorter

One of the largest portfolios of reagents




**BD**

**Learn more>**

# Assessing the causal association between 25-hydroxyvitamin D and the risk of oral and oropharyngeal cancer using Mendelian randomization.

## Authors

Tom Dudding<sup>1,2,3</sup>   
Mattias Johansson<sup>4</sup>  
Steven J Thomas<sup>3</sup>  
Paul Brennan<sup>4</sup>  
Richard M Martin<sup>1,2,5</sup>  
Nicholas J Timpson<sup>1,2,5</sup>

## Affiliations:

1.MRC Integrative Epidemiology Unit, University of Bristol, UK. 2.Bristol Medical School; Population Health Sciences, University of Bristol, UK. 3.Bristol Dental School, University of Bristol, UK. 4.International Agency for Research on Cancer, Section of Genetics - Genetic Epidemiology Group, Lyon, FR. 5.National Institute for Health Research (NIHR) Bristol Biomedical Research Centre, University Hospitals Bristol NHS Foundation Trust and the University of Bristol, Bristol, UK

**Short title:** Vitamin D – oral and oropharyngeal cancer

## Corresponding author:

Nicholas Timpson

**Address:** MRC Integrative Epidemiology Unit (IEU), Bristol Medical School: Population Health Sciences, Oakfield House, Oakfield Grove, University of Bristol, BS8 2BN

**Telephone:** +44 117 3310131

**E-mail:** [n.j.timpson@bristol.ac.uk](mailto:n.j.timpson@bristol.ac.uk)

## Key words:

Oral cancer  
Oropharyngeal cancer  
25-hydroxyvitamin D  
Mendelian randomization

## Abbreviations:

1,25OH<sub>2</sub>D = 1,25 di-hydroxyvitamin D  
25OHD = 25 hydroxyvitamin D  
CI = Confidence interval  
CEU ancestry = Northern Europeans from Utah  
GAME-ON = Genetic Associations and Mechanisms in Oncology  
GWAS = Genome wide association study  
HNC = Head and neck cancer  
MAF = Minor Allele Frequency  
MR = Mendelian randomization  
OR = Odds ratio  
QC = Quality control  
RCT = Randomized controlled trial  
SNP = Single nucleotide polymorphism

**Category:** Cancer Epidemiology

## Novelty and Impact

This is the first study to use Mendelian randomization to aid causal inference when assessing the association between 25-hydroxyvitamin D (25OHD) and oral and oropharyngeal cancers. It complements work recently published in the BMJ using similar methods to show little evidence for

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1002/ijc.31377

an association between circulating vitamin D and seven other cancer types. It is also the first Mendelian randomization study to use a genetic instrument for 25OHD that includes the low frequency variant in *CYP2R1* identified in a recent GWAS of 25OHD. This study suggests that supplementation with 25OHD is unlikely to effectively reduce the risk of oral and oropharyngeal cancer.

Accepted Article

## Abstract

Circulating 25-hydroxyvitamin D (25OHD) is an appealing potential intervention for cancer risk and has been associated with oral and oropharyngeal cancer risk but evidence is inconsistent. The availability of genetic variants, uncorrelated with known confounders, but predictive of 25OHD and genetic data in a large oral and oropharyngeal cancer collaboration aids causal inference when assessing this association.

5,133 oral and oropharyngeal cancer cases and 5,984 controls with genetic data were included in the study. Participants were based in Europe, North America and South America and were part of the Genetic Associations and Mechanisms in Oncology (GAME-ON) Network. Five genetic variants reliably associated with circulating 25OHD were used to create a relative genetic measure of 25OHD. In the absence of measured 25OHD, two-sample Mendelian randomization using individual level outcome data was used to estimate causal odds ratios (OR) for cancer case status per standard deviation increase in log25OHD. Analyses were replicated in an independent population based cohort (UK Biobank).

In the GAME-ON study there was little evidence of a causal association between circulating 25OHD and oral cancer (OR=0.86 [0.68;1.09],  $p=0.22$ ), oropharyngeal cancer (OR=1.28 [0.72;2.26],  $p=0.40$ ) or when sites were combined (OR=1.01 [0.74;1.40],  $p=0.93$ ). Replication in UK Biobank and pooled estimates produced similar results.

This study suggests that a clinically relevant protective effect of 25OHD on oral and oropharyngeal cancer risk is unlikely and supplementation of the general population with 25OHD is unlikely to be beneficial in preventing these cancers.



## Introduction

Each year there are approximately 300,000 new oral cavity and 230,000 new oropharyngeal cancers worldwide.<sup>1</sup> Squamous cell cancers of the head and neck are heterogeneous and oral and oropharyngeal cancers have differing aetiology. Alcohol and tobacco use are the major risk factors for oral and oropharyngeal cancers,<sup>2,3</sup> explaining approximately 65-70% of the population attributable risk.<sup>4</sup> Infection with human papillomavirus (HPV) has emerged as an important risk factor, particularly in oropharyngeal cancers.<sup>5</sup> The disease burden of these cancers remains high despite reductions in the prevalence of tobacco and alcohol use. Oropharyngeal cancer incidence continues to increase in the UK despite no increase in HPV infection in the past 10 years;<sup>6</sup> therefore, identification of other modifiable risk factors remains important.

Vitamin D is predominantly synthesised by exposure to ultra-violet light. The inactive metabolite (25-hydroxyvitamin D (25OHD)) is the most commonly clinically measured and supplemented form and is correlated with the active form, 1,25-dihydroxyvitamin D (1,25OH<sub>2</sub>D). An endocrine feedback mechanism controls activation of 25OHD to 1,25OH<sub>2</sub>D in the kidneys which is important for calcium and phosphate balance.<sup>7</sup> Evidence from mechanistic studies suggests a protective role of 1,25OH<sub>2</sub>D on cancer incidence and progression<sup>8,9</sup> and this may act by the same endocrine mechanism or via paracrine effects at the cancer site. It is not clear if higher circulating 25OHD is itself protective against cancer or whether other factors such as the ratio of 25OHD to 1,25OH<sub>2</sub>D at the cancer site add further complexity.<sup>10</sup>

Evidence for a role of vitamin D in head and neck cancers (HNC) is contradictory. In an observational study of tobacco-related tumours, participants with 50% lower circulating 25OHD were more likely to develop cancers of the head and neck (HNC) after accounting for available potential confounders (Hazard ratio = 1.44 (95% confidence interval (CI): 1.19, 1.73)).<sup>11</sup> A recent prospective European cohort study also found strong evidence for an inverse association between circulating 25OHD on HNC risk (Odds Ratio (OR) per doubling of circulating 25OHD concentration = 0.69 [95% CI: 0.56, 0.87],  $p=9 \times 10^{-4}$ ).<sup>12</sup> However, Arem *et al.* (2011)<sup>13</sup> and Skaaby *et al.* (2014)<sup>14</sup> found no convincing evidence of a protective effect of higher 25OHD levels on HNC.

Well-designed randomized studies of 25OHD supplementation are required to identify whether there is a true causal link with cancer incidence. However, randomized controlled trials (RCTs) are not well suited to investigate the effect of 25OHD supplementation on rare diseases. Long follow-up duration, non-adherence to intervention or control regime and an unknown optimal supplementation dose limit such trials<sup>15</sup> meaning they are prohibitively expensive and lack power to detect relatively small but potentially clinically important effects.

In the context of this lack of evidence in the literature, Mendelian randomization (MR) can help assess whether there is a causal association between 25OHD and oral and oropharyngeal cancer risk by utilising genetic data, even in studies where 25OHD itself has not been measured. MR uses genetic variants known to be reliably associated with a risk factor of interest (e.g. 25OHD) to derive estimates of the causal effect of that risk factor on health outcomes (e.g. oral cavity and oropharyngeal cancer risk).<sup>16-18</sup> Other studies have used these methods to provide evidence for causal associations of 25OHD with health outcomes including other cancers<sup>19</sup> and multiple sclerosis.<sup>20</sup>

This study aims to assess the causal association between vitamin D and oral and oropharyngeal cancer risk and estimate the size of any effect using Mendelian randomization.

## Methods



## Participants and genotyping

### GAME-ON

The study comprised 6,034 head and neck cancer cases and 6,585 controls from studies which were part of the Genetic Associations and Mechanisms in Oncology (GAME-ON) Network. Cases and controls from 10 different case-control studies were included, as well as from the European cohort study (EPIC) and cases from a UK case series (Head and Neck 5000). Details of the studies included have been described previously.<sup>21</sup> Informed consent was obtained for all participants and studies were approved by respective institutional review boards. For this study cancer cases of interest comprised the following ICD codes: oral cavity (C02.0–C02.9, C03.0–C03.9, C04.0–C04.9, C05.0–C06.9) and oropharynx (C01.9, C02.4, C09.0–C10.9). DNA extraction, genotyping, quality control and imputation has been described previously.<sup>21</sup> The study population included participants from Europe, North America and South America. To reduce the effect of heterogeneity across these regions only participants with >70% CEU ancestry were included in analyses (n=11,117). These consisted of 5,133 cases (oral n=2,700, oropharyngeal n=2,433) and 5,984 controls (**Table 1**).

### SNP selection and validation

Common genetic variants have been identified in Genome Wide Association Studies (GWAS)<sup>22,23</sup> of 25OHD, only variants that passed a genome wide association threshold ( $p < 5 \times 10^{-8}$ ) and had been replicated were selected. These genetic proxies or instruments are located in or near four 25OHD related genes: Group-Specific Component (GC), Cytochrome P450 family 2, subfamily R, polypeptide 1 (*CYP2R1*), 7-Dehydrocholesterol Reductase (*DHCR7*) and Cytochrome P450, family 24, polypeptide 1 (*CYP24A1*). A recent GWAS of 25OHD, a meta-analysis of 19 studies totalling 42,274 individuals with European ancestry, identified a low frequency variant in *CYP2R1* (rs117913124), independent of the common variant in the same gene.<sup>24</sup> Effect estimates for all SNP associations with serum 25OHD concentrations were taken from this GWAS. To assess SNP independence, SNPs on the same chromosome were assessed for linkage disequilibrium (LD) using SNAP (<http://archive.broadinstitute.org/mpg/snap/ldsearchpw.php>) in Europeans from the 1000 Genomes (Pilot 1) project. Five SNPs were identified for 25OHD from GWAS (**Table 2**). rs4588 in GC, rs10741657 in *CYP2R1* and rs6013897 in *CYP24A1* have been used in genetic instruments for 25OHD in MR studies before. rs4423214 near *DHCR7* is in perfect LD with rs12785878, identified in the original SUNLIGHT consortium 25OHD GWAS.<sup>23</sup> Finally, rs116970203 near *CYP2R1* is a low frequency variant identified in the most recent GWAS of 25OHD and has been shown to act independently of rs10741657.<sup>24</sup>

In both studies, to correct for within region population structure, region specific principal components were added to the logistic regression models. Effect allele frequencies of the five 25OHD genetic variants were calculated within the three GAME-ON regions and plotted in pie charts superimposed over a world map.

For each individual the five 25OHD variants were combined into a genetic risk score. To ease clinical interpretation, this score was converted from a per allele scale to the standardized log 25OHD scale by weighting by the beta coefficients derived from GWAS. Therefore, each unit increase in this variable (hereafter referred to as relative 25OHD) represented a standard deviation increase in log 25OHD.

### Assessment of potential confounders

One of the principals of MR is that as alleles are assigned randomly during gamete formation and segregation, the genetic instrument is not associated with factors that typically confound the observational association. To test this assumption, associations between the 25OHD instrument and all available potential confounders in GAME-ON were investigated. Sex, smoking status (never, ever, current) and alcohol use (never, ever) were recorded by questionnaire in both cohorts. GAME-ON sub-study and study country were also considered potential confounders.

Associations of the 25OHD instrument with strata of potential confounders was examined in boxplots of range and means  $\pm$  one standard deviation. Tests of associations were made using

linear regression with categorical variables as factor variables, the overall  $p$ -value for the model is shown on the boxplots.

### Replication dataset

UK Biobank is a population-based health research resource consisting of approximately 500,000 people, aged between 38 years and 73 years, who were recruited between the years 2006 and 2010 from across the UK.<sup>25</sup> Age on 31 December 2010 (coinciding with the approximate end of baseline data collection) was used as the age variable and not age at cancer diagnosis. Prevalent and incident oral and oropharyngeal cancers (ICD10 codes matched to GAME-ON) were identified from linked cancer registry data. To reduce selection bias in allocating controls, all remaining participants, after removing those participants with other HNC, were used as controls. A full description of the study design, participants and quality control (QC) methods have been described in detail previously.<sup>26</sup> UK Biobank received ethical approval from the Research Ethics Committee (REC reference for UK Biobank is 11/NW/0382).<sup>27</sup> In UK Biobank analyses were restricted to individuals of white British ancestry who self-report as “White British” and who have very similar ancestral backgrounds according to the principal component analysis, as described by Bycroft.<sup>28</sup> The full data release contains the cohort of successfully genotyped samples ( $n=488,377$ ). Any impact of genotyping array was investigated and did not meaningfully affect results (**Supplementary text**). Pre-imputation QC, phasing and imputation are described elsewhere.<sup>28</sup> Individuals with sex-mismatch or sex-chromosome aneuploidy ( $n=814$ ) and related individuals ( $n=79,448$ ) were excluded from the analysis (**Supplementary text**). After removing related and non-White British participants, there were 337,108 eligible participants, consisting of 585 cases (oral  $n=294$ , oropharyngeal  $n=291$ ) and 336,523 controls (**Supplementary table 1**).

### Statistical analyses

#### MR analyses

In a method analogous to using a genetic risk score,<sup>29</sup> relative 25OHD was used in logistic regression models to estimate a causal OR for a standard deviation increase in log 25OHD. The standard error of this estimate was corrected to account for the imprecision of the beta coefficients used to generate the relative 25OHD estimates using a bootstrap technique (**Supplementary text**).

#### GAME-ON

MR analyses were performed within each geographic region of the GAME-ON consortium accounting for age, sex and the first 15 population specific principal components. As there were likely to be differences between the causal estimates across regions, a random-effects meta-analysis was used to combine GAME-ON causal estimates using the R package ‘*meta*’.<sup>30</sup> Heterogeneity between study populations was assessed using  $I^2$ .<sup>31</sup> Meta analyses were repeated excluding geographical regions that were outliers to assess for their influence on the overall estimate.

#### Replication in UK Biobank and meta-analysis

MR analyses were repeated in UK Biobank. GAME-ON and UK Biobank estimates were compared and then combined using meta-analysis. This employed a random-effects model and was conducted using the ‘*meta*’ package in R.<sup>30</sup> Heterogeneity between studies was assessed using  $I^2$ .<sup>31</sup>

#### Sensitivity analyses

To assess whether any single SNP was driving a causal estimate, which could be driven by events such as horizontal pleiotropy which invalidate the MR assumptions, a leave-one out method was applied. This repeats the analysis sequentially removing one SNP from the genetic instrument used to derive relative 25OHD. To further demonstrate this, where a SNP overtly influences an association, each possible genetic instrument from combinations of the five 25OHD SNPs was used to estimate relative 25OHD and thereafter a causal OR. These OR were plotted as histograms highlighting those that include the highly influential SNP to see if there was an obvious grouping of these estimates.



## Power

The 25OHD genetic variants have been reported to explain 3 to 5% of the variance in 25OHD.<sup>22,23,32</sup> Given this, in the GAME-ON study for oral ( $n = 2,700$  cases, 5984 controls) or oropharyngeal sites ( $n = 2,433$  cases, 5984 controls) alone there was adequate power to detect an OR of 0.67 to 0.74 per standard deviation increase in log 25OHD with power of 0.8 and an alpha of 0.05. For all sites combined ( $n = 5,133$  cases, 5984 controls) the study was powered to detect an OR of between 0.74 and 0.79 (<http://cnsngenomics.com/shiny/mRnd/>).<sup>33</sup>

## Results

Of the five SNPs included in the analyses, rs4588 and rs10741657 were directly genotyped in both the GAME-ON and UK Biobank studies. All other SNPs were imputed and the imputation quality metric (Minimac3  $r^2$  for GAME-ON; SNPTEST info score for UK Biobank) was greater than 0.98, except rs6013897 in the GAME-ON study which had an  $r^2$  of 0.77 (**Table 2, Supplementary table 2**). No two SNPs on the same chromosome had an LD  $r^2$  greater than 0.01. Allele frequencies were similar across all three regions (**Supplementary figure 1**).

In the GAME-ON study there was no strong evidence of association of the 25OHD genetic instrument with sex, smoking status or drinking status after taking multiple testing into account. There was evidence for a difference in the 25OHD genetic instrument across regions and sub-studies (**Figure 1**). The region showing deviation from the others was South America and any differences across confounders were no longer seen when the South American region was excluded from analyses (**Supplementary figure 2 A**). Within the South American region alone there was no strong evidence for associations between the 25OHD genetic instrument and confounders (**Supplementary figure 2 B**). In UK Biobank there was no evidence for a difference in relative 25OHD across strata of sex, smoking status or drinking status (**Supplementary figure 3**).

In GAME-ON, for oral and oropharyngeal cancer there was little evidence for a causal link with relative 25OHD when the three geographic regions were meta-analysed OR = 0.86 [0.68, 1.09],  $p=0.22$  and OR = 1.28 [0.72; 2.26],  $p=0.40$  respectively. For oral sites, estimates were homogenous ( $I^2=0\%$  [0, 87],  $p_{het}=0.44$ ) but due to only having three estimates to meta-analyse (one for each geographic region), power to detect heterogeneity between regions was low. Despite this there was evidence of heterogeneity between regions for the oropharyngeal sites ( $I^2=74\%$  [16, 92],  $p_{het}=0.02$ ). This was also reflected in all sites ( $I^2=55\%$  [0, 87],  $p_{het}=0.10$ ) and was most likely driven by the strong risk increasing association with oropharyngeal cancer identified in the South American region (OR = 3.91 [1.64; 9.29],  $p<0.01$ ) compared to null associations in both European (OR = 0.98 [0.67; 1.44],  $p=0.94$ ) and North American regions (OR = 0.92 [0.64; 1.34],  $p=0.68$ ) (**Figure 2**).

When replicated in UK Biobank, estimates for oral cancer (OR=0.86 [0.50; 1.51],  $p=0.61$ ) were very similar with those from the GAME-ON study. For oropharyngeal sites (OR=0.85 [0.49; 1.47],  $p=0.56$ ) point estimates differed between GAME-ON and UK Biobank although confidence intervals overlapped. When the two studies were meta-analysed using random-effects there was no convincing evidence of a causal association with either cancer or when all sites were analysed (**Supplementary figure 4**).

## Sensitivity analyses

In leave one out analyses, sequentially omitting each of the five SNPs provided similar causal estimates for any site within the European and North American regions of GAME-ON. In the South American region of the GAME-ON study, the risk increasing association between 25OHD and oropharyngeal cancer was most strongly influenced by the rs4588 variant (**Supplementary figure 5**). In analyses using an instrument with all possible combinations of the 25OHD SNPs, even with the rs4588 variant removed there was still a risk increasing association for oropharyngeal cancer in the South American region (**Supplementary figure 6**). Removing the South American region from

the oropharyngeal analyses reduced heterogeneity between the GAME-ON studies (OR = 0.95 [0.73, 1.25],  $p=0.73$ ;  $I^2 = 0$ ,  $p_{het}=0.82$ ) (**Supplementary figure 7**).

## Discussion

This study set out to investigate the causal association between 25OHD and oral and oropharyngeal cancer risk. Given the lack of 25OHD measures in large scale oral and oropharyngeal cancer studies a two-sample MR approach,<sup>34</sup> that allowed genetic variants to be used as largely unconfounded measures of 25OHD, was used. No strong evidence of a causal association was identified with oral or oropharyngeal cancer. Given power calculations predict an OR smaller than 0.74 could be detected, if present any potential true effect is likely to be smaller than this and thus threaten the clinical relevance of the association.

These findings mirror a recently published MR study of colorectal, breast, prostate, ovarian, lung and pancreatic cancer, and neuroblastoma that did not support a causal role of 25OHD in the risk of these cancers<sup>35</sup>. Similar to the inference from this work on oral and pharyngeal cancer, the authors of the previous study state that although they are unable to rule out clinically relevant effects of small magnitude, their study combined with previous literature, “provide evidence that population-wide screening for vitamin D deficiency and subsequent widespread vitamin D supplementation should not currently be recommended as a strategy for primary cancer prevention”. Recent RCTs examining the effect of vitamin D on cancer incidence do not show strong evidence of a protective effect, though these studies are limited to older females.<sup>36,37</sup> Contrary to this, Ong *et al.* (2016)<sup>19</sup> did demonstrate a protective role for 25OHD in ovarian cancer in a well powered MR study. Despite the body of evidence that does not support a protective effect of 25OHD on cancer incidence there is evidence to support the hypothesis that vitamin D has an effect on cancer progression.<sup>8,12</sup>

The genetic instrument used in this study specifically proxies average total (free and bound) circulating 25OHD and does not necessarily predict the concentrations of free 25OHD available at the tissues or concentrations of free or bound 1,25OH<sub>2</sub>D in circulation or at the tissue level. Potential anti-cancer effects of vitamin D are via free 1,25OH<sub>2</sub>D interacting with the vitamin D receptor within tissues resulting in reduced angiogenesis, metastasis, cell invasion, inflammation, and proliferation, as well as upregulation of apoptosis.<sup>8</sup> Levels of 1,25OH<sub>2</sub>D are assumed to be correlated with circulating 25OHD, but this is not necessarily the case. Although less likely than a null association between 25OHD and oral and oropharyngeal cancer risk, the instrument being invalid for 25OHD could explain the disparity between the results of this study and previous findings showing a protective effect of 25OHD. However, as the genetic variants used here as instruments for 25OHD have robust associations with 25OHD and are in genes known to effect this metabolite they are likely to be valid for the purposes of MR.

To be valid, the genetic instrument must proxy circulating 25OHD without affecting cancer risk through other casual pathways (violation of this assumption referred to as horizontal pleiotropy). The SNPs used in this instrument are in genes with known effects on vitamin D pathways and have been consistently associated with circulating levels of 25OHD in GWA<sup>22,23</sup> and MR studies,<sup>38</sup> reducing the chance of horizontal pleiotropy. Sensitivity analyses such as MR Egger<sup>6</sup> and the weighted median method<sup>40</sup> can detect or correct for the presence of directional pleiotropy but have low power with few SNPs comprising the genetic instrument, as in this case. Here, causal estimates in the European and North American populations are largely consistent across all five SNPs, providing some evidence against the presence of strong pleiotropy that could bias findings.

In the South American population, a risk increasing association was identified between 25OHD and oropharyngeal cancer. Use of individual level data in this study allowed scrutiny of the genetic instrument-confounder independence assumption and showed that the 25OHD instrument generally held up to this assumption however there was some evidence the genetic instrument was associated with region and sub-study, driven by the South American region. Given the above points, it is unlikely then that pleiotropy or confounding can explain the large risk increasing association seen for oropharyngeal cancer in the South American region. However, as this risk increasing

association is inconsistent with observational effect directions and MR estimates from other geographic regions, it is unlikely to be truly causal: it most likely represents structure within this population that is by chance associated with oropharyngeal cancer risk. To assess the impact of this potentially biased estimate on the overall causal OR, the analyses were repeated with the South American individuals removed, resulting in more consistent estimates.

## Conclusion

This study does not support the observational association between 25OHD and oral cancer risk and is consistent with evidence that a causal, clinically relevant protective effect of 25OHD on oropharyngeal cancer risk is unlikely. The effect of 25OHD on oral and oropharyngeal cancer progression was not assessed here and requires further investigation.

The genetic instrument for 25OHD is a good proxy for circulating total 25OHD, the same metabolite that is used in supplementation and is measured in the observational studies. Despite being robustly associated with 25OHD, the genetic variants used here only explain a relatively small amount of variance in 25OHD. This means that causal effects of small magnitude cannot be ruled out. Furthermore, it cannot be determined whether 25OHD has an effect on cancer risk in individuals below a certain 25OHD concentration threshold. Where inference is likely to be useful clinically is in relation to supplementation. Supplementation increases total circulating 25OHD and is presumed to subsequently increase free 1,25OH<sub>2</sub>D availability at the tissues. The standard deviation 25OHD increase in this study will have a relative effect on total 25OHD across the whole life course. If this long-term difference in relative 25OHD does not produce a detectable effect in this study, any 25OHD causal effect is unlikely to be of a magnitude that would warrant long-term supplementation in the general population from a particular age.

## **Funding**

The MRC/University of Bristol Integrative Epidemiology Unit (IEU) is supported by the MRC and the University of Bristol (MC\_UU\_12013/1-9)

RMM, NJT, PB are supported by a Cancer Research UK Programme Grant, the Integrative Cancer Epidemiology Programme (C18281/A19169).

This research is funded by the National Institute for Health Research (NIHR) Bristol Biomedical Research Centre which is funded by the National Institute for Health Research and is a partnership between University Hospitals Bristol NHS Trust and the University of Bristol.

TD is supported by the Wellcome Trust (201237/Z/16/Z).

NJT is a Wellcome Trust Investigator (202802/Z/16/Z), is a programme lead in the MRC Integrative Epidemiology Unit (MC\_UU\_12013/3) and works within the University of Bristol NIHR Biomedical Research Centre (BRC).

## **Acknowledgments**

Quality Control filtering of the UK Biobank data was conducted by R.Mitchell, G.Hemani, T.Dudding, L.Paternoster as described in the published protocol (doi:doi.org/10.5523/bris.3074krb6t2frj29yh2b03x3wxj).

## References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 2015;136:E359–86.
2. Maier H, Dietz A, Gewelke U, Heller WD, Weidauer H. Tobacco and alcohol and the risk of head and neck cancer. *Clin Invest* 1992;70–70:320–7.
3. Freedman ND, Schatzkin A, Leitzmann MF, Hollenbeck AR, Abnet CC. Alcohol and head and neck cancer risk in a prospective study. *Br J Cancer* 2007;96:1469–74.
4. Hashibe M, Brennan P, Chuang S-C, Boccia S, Castellsague X, Chen C, Curado MP, Dal Maso L, Daudt AW, Fabianova E, Fernandez L, Wünsch-Filho V, et al. Interaction between tobacco and alcohol use and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *Cancer Epidemiol Biomarkers Prev* 2009;18:541–50.
5. Chaturvedi AK, Engels EA, Pfeiffer RM, Hernandez BY, Xiao W, Kim E, Jiang B, Goodman MT, Sibug-Saber M, Cozen W, Liu L, Lynch CF, et al. Human Papillomavirus and Rising Oropharyngeal Cancer Incidence in the United States. *J Clin Oncol* 2011;29:4294–301.
6. Schache AG, Powell NG, Cuschieri KS, Robinson M, Leary S, Mehanna H, Rapozo D, Long A, Cubie H, Junor E, Monaghan H, Harrington KJ, et al. HPV-related oropharynx cancer in the United Kingdom: An evolution in the understanding of disease etiology. *Cancer Res* 2016;76:6598–606.
7. Holick MF. Vitamin D and bone health. *J Nutr* 1996;126:1159S–64S.
8. Feldman D, Krishnan A V., Swami S, Giovannucci E, Feldman BJ. The role of vitamin D in reducing cancer risk and progression. *Nat Rev Cancer* 2014;14:342–57.
9. Deeb KK, Trump DL, Johnson CS. Vitamin D signalling pathways in cancer: potential for anticancer therapeutics. *Nat Rev Cancer* 2007;7:684–700.
10. Tan V, Biernack K, Dudding T, Bonilla C, Gilbert R, Kaplan R, Qibin Q, Teumer A, Martin RM, Perks C, Timpson NJ, Holly JMP. The association between circulating Vitamin D and IGFBP-3: observational and causal estimates from independent sources. Under review.
11. Afzal S, Bojesen SE, Nordestgaard BG. Low Plasma 25-Hydroxyvitamin D and Risk of Tobacco-Related Cancer. *Clin Chem* 2013;59.
12. Fanidi A, Muller DC, Midttun Ø, Ueland PM, Vollset SE, Relton C, Vineis P, Weiderpass E, Skeie G, Brustad M, Palli D, Tumino R, et al. Circulating vitamin D in relation to cancer incidence and survival of the head and neck and oesophagus in the EPIC cohort. *Sci Rep* 2016;6:36017.
13. Arem H, Weinstein SJ, Horst RL, Virtamo J, Yu K, Albanes D, Abnet CC. Serum 25-Hydroxyvitamin D and Risk of Oropharynx and Larynx Cancers in Finnish Men. *Cancer Epidemiol Biomarkers Prev* 2011;20:1178–84.
14. Skaaby T, Husemoen LLN, Thuesen BH, Pisinger C, Jørgensen T, Roswall N, Larsen SC, Linneberg A. Prospective population-based study of the association between serum 25-hydroxyvitamin-D levels and the incidence of specific types of cancer. *Cancer Epidemiol Biomarkers Prev* 2014;23:1220–9.
15. IARC. Vitamin D and Cancer. 2008.
16. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. Mendelian randomization:



- using genes as instruments for making causal inferences in epidemiology. *Stat Med* 2008;27:1133–63.
17. Davey Smith G, Ebrahim S. “Mendelian randomization”: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;32:1–22.
  18. Hingorani A, Humphries S. Nature’s randomised trials. *Lancet* 2005;366:1906–8.
  19. Ong J-S, Cuellar-Partida G, Lu Y, Hein A, Burghaus S, Beckmann MW, Vanderstichele A, Doherty JA, Rossing MA, Rudolph A, Wang-Gohrke S, Goodman MT, et al. Association of vitamin D levels and risk of ovarian cancer: a Mendelian randomization study. *Int J Epidemiol* 2016;1619–30.
  20. Mokry LE, Ross S, Ahmad OS, Forgetta V, Smith GD, Leong A, Greenwood CMT, Thanassoulis G, Richards JB. Vitamin D and Risk of Multiple Sclerosis: A Mendelian Randomization Study. *PLoS Med* 2015;12:e1001866.
  21. Lesseur C, Diergaarde B, Olshan AF, Wünsch-Filho V, Ness AR, Liu G, Lacko M, Eluf-Neto J, Franceschi S, Laggiou P, Macfarlane GJ, Richiardi L, et al. Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer. *Nat Genet* 2016;48:1544–50.
  22. Ahn J, Yu K, Stolzenberg-Solomon R, Claire Simon K, McCullough ML, Gallicchio L, Jacobs EJ, Ascherio A, Helzlsouer K, Jacobs KB, Li Q, Weinstein SJ, et al. Genome-wide association study of circulating vitamin D levels. *Hum Mol Genet* 2010;19:2739–45.
  23. Wang TJ, Zhang F, Richards JB, Kestenbaum B, van Meurs JB, Berry D, Kiel DP, Streeten EA, Ohlsson C, Koller DL, Peltonen L, Cooper JD, et al. Common genetic determinants of vitamin D insufficiency: a genome-wide association study. *Lancet* 2010;376:180–8.
  24. Manousaki D, Dudding T, Haworth S, Hsu Y-H, Liu C-T, Medina-Gómez C, Voortman T, van der Velde N, Melhus H, Robinson-Cohen C, Cousminer DL, Nethander M, et al. Low-Frequency Synonymous Coding Variation in CYP2R1 Has Large Effects on Vitamin D Levels and Risk of Multiple Sclerosis. *Am J Hum Genet* 2017;29:98–104.
  25. Allen NE, Sudlow C, Peakman T, Collins R. UK Biobank Data: Come and Get It. *Sci Transl Med* 2014;6:224ed4–224ed4.
  26. Collins R. What makes UK Biobank special? *Lancet* 2012;379:1173–4.
  27. Mitchell R, Hemani G, Dudding T, Paternoster L. UK Biobank Genetic Data: MRC-IEU Quality Control, Version 1. University of Bristol, 2017.
  28. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O’Connell J, Cortes A, Welsh S, et al. Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv* 2017;166298.
  29. Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. *Int J Epidemiol* 2013;42:1134–44.
  30. Schwarzer G. meta: An R package for meta-analysis. *R News* 2007;7:40–5.
  31. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ Br Med J* 2003;327:557–60.
  32. Hiraki LT, Major JM, Chen C, Cornelis MC, Hunter DJ, Rimm EB, Simon KC, Weinstein SJ, Purdue MP, Yu K, Albanes D, Kraft P. Exploring the Genetic Architecture of Circulating 25-Hydroxyvitamin D. *Genet Epidemiol* 2013;37:92–8.
  33. Brion MJA, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol* 2013;42:1497–501.
  34. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG, EPIC- InterAct Consortium E-I. Using published data in Mendelian randomization: a blueprint for efficient

- identification of causal risk factors. *Eur J Epidemiol* 2015;30:543–52.
35. Dimitrakopoulou VI, Tsilidis KK, Haycock PC, Dimou NL, Al-Dabhani K, Martin RM, Lewis SJ, Gunter MJ, Mondul A, Shui IM, Theodoratou E, Nimpitsch K, et al. Circulating vitamin D concentration and risk of seven cancers: Mendelian randomisation study. *BMJ* 2017;359:j4761.
  36. Hill KB, Chadwick BL, Freeman R, O'Sullivan I, Murray JJ. Adult Dental Health Survey 2009: Relationships between dental attendance patterns, oral health behaviour and the current barriers to dental care. *Br Dent J* 2013;214:25–32.
  37. Lappe J, Watson P, Travers-Gustafson D, Recker R, Garland C, Gorham E, Baggerly K, McDonnell SL. Effect of Vitamin D and Calcium Supplementation on Cancer Incidence in Older Women. *JAMA* 2017;317:1234.
  38. Vimalaswaran KS, Berry DJ, Lu C, Tikkanen E, Pilz S, Hiraki LT, Cooper JD, Dastani Z, Li R, Houston DK, Wood AR, Michaëlsson K, et al. Causal Relationship between Obesity and Vitamin D Status: Bi-Directional Mendelian Randomization Analysis of Multiple Cohorts. *PLoS Med* 2013;10.
  39. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015;44:512–25.
  40. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol* 2016;40:304–14.

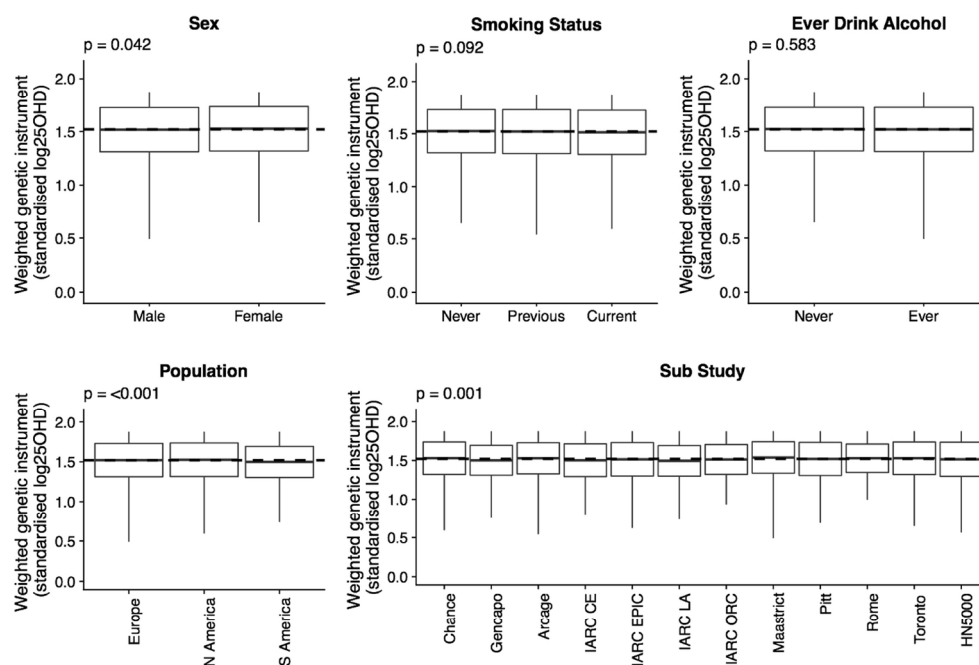
**Table 1: GAME-ON participant summaries**

	All n (%) N=11117	Case n (%) N=5133	Control n (%) N=5984
Site			
Oral		2700 (52.6)	
Oropharyngeal		2433 (47.4)	
Age			
=<50	2217 (19.9%)	1071 (20.9%)	1146 (19.2%)
50-<60	3443 (31.0%)	1696 (33.0%)	1747 (29.2%)
60-<70	3348 (30.1%)	1504 (29.3%)	1844 (30.8%)
>=70	2108 (19.0%)	861 (16.8%)	1247 (20.8%)
Missing	1 (0.01%)	1 (0.02%)	0 (0.00%)
Sex			
Male	7680 (69.1%)	3798 (74.0%)	3882 (64.9%)
Female	3437 (30.9%)	1335 (26.0%)	2102 (35.1%)
Smoking status			
Never	3302 (29.7%)	1002 (19.5%)	2300 (38.4%)
Previous	3655 (32.9%)	1590 (31.0%)	2065 (34.5%)
Current	3300 (29.7%)	2019 (39.3%)	1281 (21.4%)
Missing	860 (7.74%)	522 (10.2%)	338 (5.65%)
Alcohol use			
Never	1825 (16.4%)	767 (14.9%)	1058 (17.7%)
Ever	8422 (75.8%)	4030 (78.5%)	4392 (73.4%)
Missing	870 (7.83%)	336 (6.55%)	534 (8.92%)
Geographic region			
Europe	5251 (47.2%)	2323 (45.3%)	2928 (48.9%)
North America	4583 (41.2%)	2254 (43.9%)	2329 (38.9%)
South America	1283 (11.5%)	556 (10.8%)	727 (12.1%)

Table 2: 25-Hydroxyvitamin D genetic variant details

RSID	Chromosome	Position	Gene	Effect Allele	Other Allele	Vitamin D GWAS				GAME ON	
						EAF	Beta	SE	P value	Rsq (minimac3)	EAF
rs4588	4	72618323	GC	G	T	0.717	0.2469	0.0070	1.68E-263	Genotyped	0.722
rs116970203	11	14876718	PDE3B	G	A	0.975	0.4323	0.0209	2.29E-90	0.98	0.978
rs4423214	11	71173254	DHCR7	T	C	0.697	0.0998	0.0073	1.39E-40	0.99	0.718
rs10741657	11	14914878	CYP2R1	A	G	0.415	0.0938	0.0065	8.76E-45	Genotyped	0.383
rs6013897	20	52742479	CYP24A1	T	A	0.791	0.0658	0.0080	9.06E-16	0.77	0.781

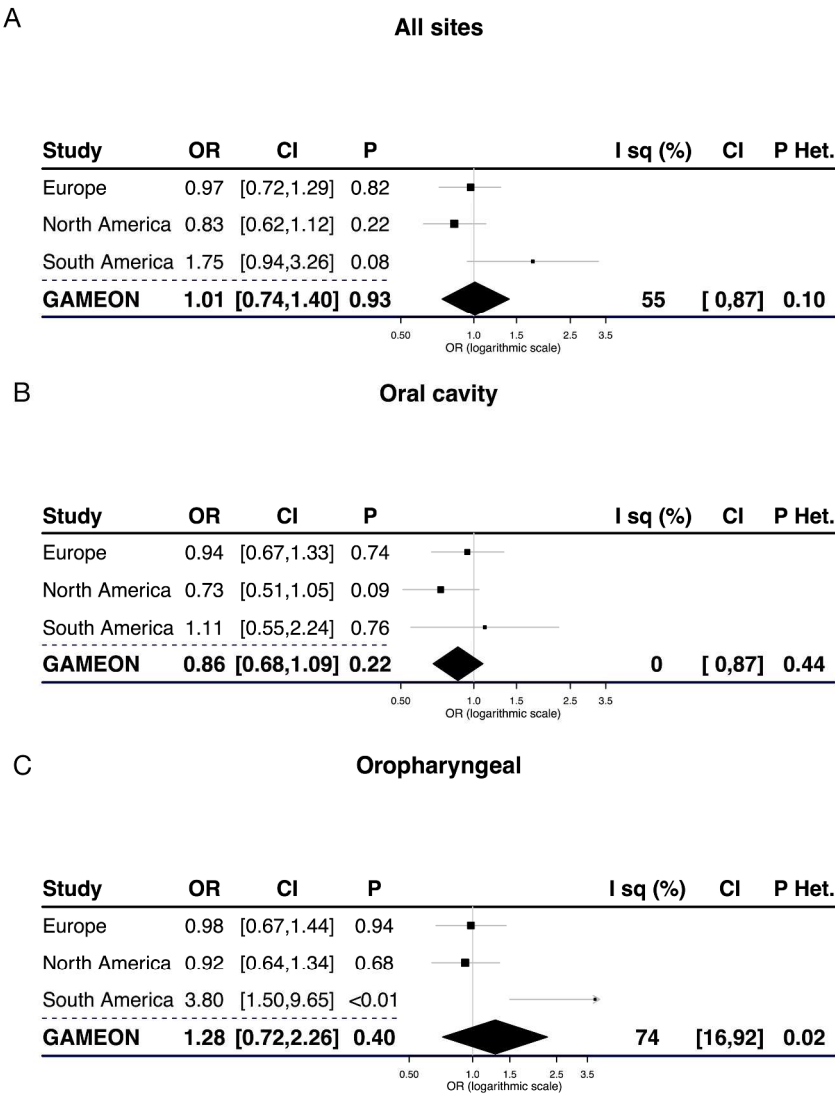
EAF = effect allele frequency, Rsq = R squared, SE = standard error



**Figure 1: Mean (+/- standard deviation) and range of 25-hydroxyvitamin D weighted genetic instrument across strata of potential confounders for GAMEON.** Overall mean 25-hydroxyvitamin D weighted genetic instrument for those included in plot indicated by dashed line. 25OHD = 25 hydroxyvitamin D, p=p value for linear model

144x105mm (300 x 300 DPI)





**Figure 2: Odds ratio (95% CI) for developing cancer for a SD increase in 25-hydroxyvitamin D for all sites (A), oral sites (B) and oropharyngeal sites (C) in Mendelian randomization analysis.** CI = 95% confidence interval, OR = odds ratio, P = p value, P Het = p value for heterogeneity

139x180mm (600 x 600 DPI)